California Department of Education

Executive Office

SBE-002 (REV. 11/2017)

memo-imab-adad-apr22item01

# **MEMORANDUM**

**DATE:** April 12, 2022

**TO:** MEMBERS, State Board of Education

**FROM:** TONY THURMOND, State Superintendent of Public Instruction

**SUBJECT:** Alternate English Language Proficiency Assessments for California Operational Field Test: Standard Setting Plan.

## Summary of Key Issues

This Information Memorandum provides a summary of activities related to the February/March 2022 Alternate English Language Proficiency Assessments for California (ELPAC) Standard Setting Workshop, a virtual workshop during which California educators collaborated to develop recommended threshold scores. In May 2022, the California Department of Education (CDE) will bring proposed Alternate ELPAC threshold scores to the California State Board of Education (SBE) for approval. If the proposed threshold scores receive the SBE’s approval, Student Score Reports (SSRs) for the 2021–22 Alternate ELPAC operational field test administration will be available to local educational agencies (LEAs) in July 2022, and summary results will be released shortly thereafter.

Standard setting is an empirical process that uses a panel of experts to recommend threshold scores for reporting performance levels. During standard setting, educators use range performance level descriptors (PLDs) to create the threshold student definitions. Using a consensus building process, panelists make judgments that define what a student “can do” at the entry to each performance level. The threshold scores are then used to include performance level information with test results to students and parents.

## Alternate ELPAC Standard Setting Process

The Alternate ELPAC Standard Setting Workshop was conducted by testing contractor ETS on February 22 through March 3, 2022, and observed by CDE staff. ETS workshop facilitators provided training and practice in the standard setting process and the use of the PLDs as a tool in the development of panel-recommended threshold scores.

The 44 educators who served as panelists formed six panels, one for each grade or grade-span assessment. The panelists were a diverse group, all of whom who were familiar with the *2012 California English Language Development Standards* (*2012 ELD Standards*) and ELD Connectors and experienced in teaching EL students with the most significant cognitive disabilities. Nine educators participated on two different grade level panels.

Of the three performance levels for the Alternate ELPAC—Level 1, novice English learner (EL); Level 2, intermediate EL; and Level 3, fluent English proficient—Levels 2 and 3 were used at the Standard Setting Workshop. The panelists developed descriptions of what students can do at the beginning of Levels 2 and 3. Using those descriptions, the panelists made judgments about each item, and those judgments resulted in the setting of threshold scores at Level 2 and Level 3.

**Approaches Used**

For the Standard Setting Workshop, ETS adhered to the guidelines and best practices proposed in the standard setting literature, using two widely accepted standard setting approaches—the modified Angoff method (Brandon, 2004; Hambleton & Pitoniak, 2006) and the extended Angoff method (Cizek & Bunch, 2007; Hambleton & Plake, 1995). For the Alternate ELPAC one-point items, the modified Angoff method was used; for the two-point items, the extended Angoff method was used. The process was conducted over two rounds of judgments and included feedback and discussion between the rounds. Panel-recommended threshold scores were based on the panel average.

After the second round of panel judgments, representatives from each of the six grade and grade-span panels assembled for a vertical articulation meeting, during which the standard setting process was discussed. During vertical articulation, the representatives reviewed how the six panels (1) identified through discussion what was being measured by the Alternate ELPAC; (2) developed threshold student definitions; and (3) incorporated feedback when making independent judgments. The purpose of this meeting was to review the work of the panels for all grades and grade spans; no discrepancies in the process were noted. The final panel-recommended threshold scores for Level 2 and Level 3 were based on the bank of items used to assemble both the Initial Alternate ELPAC and Summative Alternate ELPAC and could, therefore, be applied to both assessments.

Following the February/March 2022 Alternate ELPAC Standard Setting Workshop, ETS reported the standard setting panels’ recommendations to the CDE for review. Psychometricians from the CDE and selected ELPAC Technical Advisory Group members completed a review of the panel recommendation judgments along with other data, which will inform the recommended State Superintendent of Public Instruction’s threshold scores that will be presented to the SBE in May 2022 for approval.

Attachment 1, “Standard Setting Plan for the Alternate English Language Proficiency Assessments for California,” provides detailed information on the standard setting process used in this Standard Setting Workshop.

## Key Activities in the Development of the Alternate ELPAC

Table 1 provides a timeline of key activities in the development of the Alternate ELPAC.

**Table 1. Timeline of Key Development Activities—Alternate ELPAC**

| **Activity** | **Date(s)** |
| --- | --- |
| SBE approved the high-level test design and general PLDs | May 2019 |
| SBE approved the test blueprint | May 2020 |
| Statewide administration of the Alternate ELPAC operational field test | November 1, 2021, through May 31, 2022 |
| ETS conducted the Alternate ELPAC Standard Setting Workshop with California educators | February 22 through March 3, 2022 |
| SBE considers for approval the Alternate ELPAC threshold scores | May 2022 |
| ETS provides Alternate ELPAC SSRs to LEAs for distribution to parents (contingent on approval of threshold scores in May)\* | July 2022  (tentative) |
| First Operational Initial Alternate ELPAC testing window | July 1, 2022, through June 30, 2023 |
| First Operational Summative Alternate ELPAC testing window | February 1 through May 31, 2023 |

\*Contingent on SBE approval of the Alternate ELPAC threshold scores in May 2022.

### **Next Steps for Alternate ELPAC Threshold Scores**

In May 2022, the CDE will bring the recommended State Superintendent of Public Instruction’s Alternate ELPAC threshold scores to the SBE for approval. If the SBE approves these recommended threshold scores, LEAs will receive electronic SSRs to send to parents/guardians and students. In addition to the SSRs, LEAs also will be able to access student results in a downloadable student-level data file in the Test Operations Management System. Once available, the statewide-, county-, LEA-, and school-level summary results by student groups will be displayed on the Test Results for California's Assessments website at <https://caaspp-elpac.cde.ca.gov/elpac/Default>.

## Attachment(s)

* Attachment 1: Standard Setting Plan for the Alternate English Language Proficiency Assessments for California (14 Pages)



**Standard Setting Plan for the Alternate English Language Proficiency Assessments for California**

**Contract #CN150012**

**Prepared for the California Department of Education by ETS**

**Presented December 23, 2021**



**Contents**

[**1. Background 3**](#_Toc89343710)

[**2. Alternate ELPAC 4**](#_Toc89343711)

[**3. Description of the Standard Setting 5**](#_Toc89343712)

[3.A. Purpose 5](#_Toc89343713)

[3.B. General Description of the Process 5](#_Toc89343714)

[3.C. Panelists 5](#_Toc89343715)

[**4. Standard Setting Materials 6**](#_Toc89343716)

[4.A. Judgment Training Materials 7](#_Toc89343717)

[4.B. Evaluation Forms 7](#_Toc89343718)

[**5. Standard Setting Process 7**](#_Toc89343719)

[5.A. Test Familiarization 8](#_Toc89343720)

[5.B. Defining Threshold Students 9](#_Toc89343721)

[**6. Standard Setting Methodology 9**](#_Toc89343722)

[6.A. Item Scoring, Judgments, and Rating Scales 10](#_Toc89343723)

[6.B. Modified and Extended Angoff Judgments 10](#_Toc89343724)

[**7. Feedback and Discussion 12**](#_Toc89343725)

[7.A. Round 1 Feedback and Discussion, and Round 2 Judgments 12](#_Toc89343726)

[7.B. Round 2 Feedback and Discussion 13](#_Toc89343727)

[7.C. Vertical Articulation 13](#_Toc89343728)

[7.D. Final Recommendations 13](#_Toc89343729)

[**8. Recommendations and Technical Report 14**](#_Toc89343730)

[**9. References 14**](#_Toc89343731)

## Background

This plan describes the process that ETS will use to conduct standard setting for the Alternate English Language Proficiency Assessments for California (ELPAC). The standard setting process will result in educator-recommended threshold scores for the Alternate ELPAC, which will be delivered to the California Department of Education (CDE). The recommendations are scheduled to be presented to the California State Board of Education (SBE) in May 2022 along with the recommendation from the State Superintendent of Public Instruction (SSPI).

In 2017–18, ETS, under the direction of the CDE, launched the ELPAC. Beginning in 2018, the CDE, together with ETS, began developing an alternate English language proficiency (ELP) assessment: the Alternate ELPAC. The term “Alternate ELPAC” includes both the initial and summative administrations of the assessment. This assessment is designed to assess the ELP of English learners with the most significant cognitive disabilities (ELSCDs) and potential ELSCDs, as determined by the individualized education program (IEP) team. The Alternate ELPAC is aligned with the [2012 California English Language Development (ELD) Standards](https://www.cde.ca.gov/sp/el/er/documents/eldstndspublication14.pdf), which were adopted by the SBE, via the ELD Connectors.

These test takers are a diverse student population in kindergarten through grade twelve, including students who are under twenty-two years of age and enrolled in grade twelve who continue to be eligible for special education and English learner (EL) program and support services. The variety of students’ language- and disability-related needs and alternate ways of communicating requires careful consideration when thinking about how to measure the ELP needed to communicate in social and academic contexts. The CDE and ETS have designed and developed a computer-based Alternate ELPAC that is intended to meet federal requirements and best supports ELSCDs in their progress toward meeting ELP.

In May 2019, the SBE approved the Alternate ELPAC general performance level descriptors (PLDs), as listed in the [Proposed High-Level Test Design for the Alternate ELPAC](https://www.cde.ca.gov/ta/tg/ep/documents/proposedhltdaltelpac.pdf). These general, or policy, PLDs convey the degree of a student’s ELP using three levels: Level 1, novice English learner; Level 2, intermediate English learner; and Level 3, fluent English proficient. In fall 2020, range PLDs were reviewed by California educators and approved by the CDE. Range PLDs further describe what students at each level know and can do. They are used in standard setting to distinguish between the meaning of each performance level at a grade level or grade span. Range PLDs are important because they define the knowledge, skills, and abilities required at each performance level for a grade level or grade span assessed. The standard setting panel will recommend threshold scores that indicate the score that must be earned for a student to reach the beginning (i.e., threshold) of the second and third performance levels. Descriptions of the process, panel composition, materials, and methodology are provided in this document.

## Alternate ELPAC

The Alternate ELPAC assesses the four domains of Listening, Reading, Speaking, and Writing. However, it does so in an integrated manner; that is, a single task type assesses multiple domains, either receptive (Listening and Reading) or expressive (Speaking and Writing). In addition, to ensure that students can fully access and participate in the Alternate ELPAC, these four domains are assessed via students’ individually preferred receptive and expressive communication modes. Individually preferred communication modes are the ways in which a student typically comprehends and expresses information. They include listening, reading, speaking, and writing as well as sign language, eye gaze, pointing, gestures, augmentative and alternative communication devices, and other alternate modes to comprehending and expressing information. The use of receptive and expressive categories on the Alternate ELPAC allows maximum flexibility for students to demonstrate ELP through the means that are most consistent with how they communicate in the classroom.

The Alternate ELPAC consists of two assessments: (1) the Initial Alternate ELPAC; and (2) the Summative Alternate ELPAC. The assessments are administered to eligible students, as determined by the IEP team. The purpose of the Alternate ELPAC is twofold:

* The Initial Alternate ELPAC will provide information to determine a student’s initial classification as an EL or as initial fluent English proficient (IFEP), for students with the most significant cognitive disabilities.
* The Summative Alternate ELPAC will provide information on annual student progress toward ELP and support decisions on whether or not a student is ready to be reclassified as fluent English proficient (RFEP), for students with the most significant cognitive disabilities.

The Alternate ELPAC aligns with the 2012 ELD Standards via the ELD Connectors. The ELD Connectors offer a reduction in the depth, breadth, and complexity of the standards, as appropriate for students with the most significant cognitive disabilities. In addition to being directed by the 2012 ELD Standards, the ELD Connectors were substantially informed by the ELP level descriptors in the Council of Chief State School Officers’ ELP Standards for English Learners with Significant Cognitive Disabilities (CCSSO, 2019).

The Initial Alternate ELPAC will be administered once to eligible students entering the California school system for the first time, in the following grade levels and grade spans:

* Kindergarten
* One
* Two
* Three through five
* Six through eight
* Nine through twelve

The Summative Alternate ELPAC will be administered annually in the spring to eligible students in the following grade levels and grade spans:

* Kindergarten
* One
* Two
* Three through five
* Six through eight
* Nine and ten
* Eleven and twelve

## Description of the Standard Setting

### **Purpose**

The Alternate ELPAC standard setting will result in recommendations for the Alternate ELPAC threshold scores, which are the minimum scores at Level 2 and Level 3 for each grade level and grade span. The CDE will review these recommendations, along with additional data; these and the SSPI recommendations are scheduled to be presented to the SBE in May 2022.

### **General Description of the Process**

The approach used in this study adheres to the guidelines and best practices recommended in the standard setting literature; specifically, the modified Angoff and extended Angoff standard setting methods (Cizek & Bunch, 2007; Hambleton & Plake, 1995) will be implemented. Panelists will make two rounds of judgments, with feedback and discussion occurring between each round. Once judgments have been made for all six grade-level and grade-span assessments, a final review across all threshold scores and assessments will take place. Representatives from each of the six panels will participate in the Vertical Articulation Meeting, which is described in the [Vertical Articulation](#_Vertical_Articulation) section of this document.

### **Panelists**

In recruiting panelists, the goal is to include a range of California educators familiar with instructing this student population. Panelists should meet the following qualifications:

* Familiarity with the ELD Connectors
* Experience in the education of students who are eligible to take the Alternate ELPAC

Further, the goal is to recruit experienced panelists who meet each of the following qualifications:

* They are engaged in the daily instruction of students with the most significant cognitive disabilities, especially ELSCDs.
* They are engaged in the daily instruction of the 2012 ELD Standards and ELD Connectors.
* They understand and represent the diverse group of students eligible to take the Alternate ELPAC.
* They have indicated that they will administer the Alternate ELPAC operational field test, between November 2021 and February 2022.

Teachers who participated in Alternate ELPAC educator meetings and those who previously agreed to participate in the standard setting scheduled during 2020–21 will be given special consideration. Additional effort will be made to include educators who have other relevant experience, such as speech-language pathologists or ELD teachers; however, to participate, those panelists are required to be trained in the administration of the Alternate ELPAC. The decision on the panelists’ selection and panel assignment will be made by the CDE.

Each panel, which should include 8 to 10 panelists from this population of educators, will attend a workshop specific to the grade-level or grade-span assignment, as shown in table 1. The workshop is planned for three and one-half days, and a vertical articulation panel, described in the [Vertical Articulation](#_Vertical_Articulation) section, will meet on March 3, after the upper-grades workshops are completed.

Table 1. Educator Panel Workshops in 2022

|  |  |
| --- | --- |
| **Panel** | **Workshop Dates** |
| Kindergarten | February 22–25 |
| Grade one | February 22–25 |
| Grade two | February 22–25 |
| Grade span three through five | February 28–March 3 |
| Grade span six through eight | February 28–March 3 |
| Grade span nine through twelve\* | February 28–March 3 |
| Vertical articulation | March 3 |

**\***The standard setting requires high school educators to recommend one set of threshold scores for the high school assessments. This is because the grade span nine and ten and grade span eleven and twelve summative assessments are based on the same standards, ELD Connectors, and blueprints and have the same general and range PLDs.

Panelists will be notified after being approved, and the Sacramento County Office of Education is assisting with panelist communication and meeting logistics. Panelists will be polled; if they have not administered the test, a training module will be included as part of their prework for the standard setting. All panelists will be required to sign a security agreement acknowledging that standard setting materials are confidential and, therefore, must be destroyed prior to the end of the workshop.

## Standard Setting Materials

The materials listed in this section will be provided to panelists for each assessment. The manner in which these materials are used is described more fully in the [Standard Setting Process](#_Standard-Setting_Process) and [Standard Setting Methodology](#_Standard-Setting_Methodology) sections. Nonsecure premeeting materials will be sent to panelists’ email addresses; other materials will be provided during the workshop when needed for each stage of the process. Materials used during the workshop will be provided using remote tools, such as the ETS Content Review Tool (CRT) and Zoom web conference software.

The following materials will be provided to panelists:

* Workshop agenda
* Alternate ELPAC range PLDs and the ELD Connectors, provided in the premeeting materials as part of preparation for the workshop
* Training materials, provided as prework (i.e., training website link and the *Directions for Administration* [*DFAs*]for the training test)
* Alternate ELPAC operational field test materials
* Provided to panelists during the workshop through the use of ETS’s CRT (All unique items administered for the assessment will be provided to the panel.)
* Provided to panelists along with an answer key, including scoring rules and rubrics, where needed
* Judgment materials
* Used during the workshop, including forms that are unique to each panelist, and provided electronically and remotely
* Standard setting evaluation survey forms: training evaluation forms and a final evaluation form, provided electronically

### **Judgment Training Materials**

The purpose of the judgment training is to familiarize the panelists with the types of judgments to be made. During the standard setting training, the panelists will make judgments on a sample of items from the Alternate ELPAC training test. The training process, which includes learning how to make judgments on online forms and for each of the judgment types, is described in the [Standard Setting Methodology](#_Standard-Setting_Methodology) section.

During the operational rounds of judgments, panelists will again enter judgments on online forms. At that time, operational judgments will be made on all unique items administered to Alternate ELPAC test takers.

### **Evaluation Forms**

Following each training round, the panelists will be asked to complete a training evaluation, during which the panelists can indicate their understanding of the judgment-making process. At the conclusion of the workshop, panelists will be asked to complete a final evaluation of the entire process.

## Standard Setting Process

The process will include information and assignments for panelists to review in advance as well as interactive work and independent work that will be completed during the workshop dates. Approximately two weeks before the workshop, panelists will receive a three-part preworkshop assignment. For the part one, panelists will prepare their devices (e.g., laptop, tablet) and software for the workshop using links and instructions received from ETS. For part two, panelists will access the Alternate ELPAC training test to refamiliarize themselves with the assessment’s item types. All panelists will have previously administered the Alternate ELPAC or will be assigned to take the test administration training. For part 3 of the preworkshop assignment, panelists will make notes on a notetaking form for reviewing the general PLDs and the rangePLDs for their assigned grade level or grade span; panelists will receive the ELD Connectors for reference. This task will be in preparation for discussion at the workshop.

Each week, the workshop will begin with a general session and will include panelists across grade-level and grade-span panels (e.g., kindergarten, grade one, and grade two in week one). The ETS standard setting director will provide an overview of the Alternate ELPAC standard setting process, including roles and responsibilities of the educators, the CDE, and ETS staff. ETS will invite general process questions as well as policy questions; the CDE will provide responses to policy questions. During the rest of the process, individual panels will work on one grade-level or grade-span assessment with a panel facilitator. Each facilitator has over 10 years of experience working with educators in standard setting. In the Zoom web conference panel rooms (hereafter referred to as panel rooms), facilitators will provide in-depth training and practice on the method for each step of the process. Panelists will complete two rounds of judgments, and feedback and discussion will take place after each round of judgment. Each panel will complete the standard setting process on one grade-level or grade-span assessment. At the end of the last panel meeting, there will be a two-hour vertical articulation meeting where representatives from each panel will review the process and recommendations across all six panels.

### **Test Familiarization**

At each panel room session, the first activity will be for the panelists to have the experience of taking the test, with an ETS assessment developer acting as a test examiner by presenting the test items to all panelists. All items administered to students in the 2021–22 operational field test will be included in the workshop. The panelists will access all materials in the same way students taking the Alternate ELPAC would access them. This activity will allow panelists to become familiar with Alternate ELPAC test items and discuss the test content before making any judgments. Panelists will be reminded that this is not an item review; rather, the purpose is to share thoughts about what the Alternate ELPAC is measuring, what they think may be particularly challenging for students, and what they think might be less difficult for students. The goal of the discussion is for panelists to share with each other their perceptions of the general difficulty, for students, of the tested material.

Once the panelists are familiar with the assessment, the group will begin discussing notes taken during the preworkshop assignment, specifically articulating the knowledge and skills necessary for students to reach performance Level 3, using the range PLDs. The initial focus will be on the draft bullets for the Level 3 threshold student that were included in the preworkshop assignment.

### **Defining Threshold Students**

Developing definitions of threshold students is a critical component of any standard setting workshop. The process to arrive at threshold student definitions will involve small-group discussions and the development of draft threshold student definitions, followed by a whole-panel discussion of the draft definitions. The goal is to develop a panel consensus of what is expected.

For the Alternate ELPAC, two definitions will be needed for two thresholds—the Level 2 threshold and the Level 3 threshold. Within each grade level and grade span, the thresholds will be applied to the Initial Alternate ELPAC and the Summative Alternate ELPAC, as the blueprints and PLDs are the same across both assessments. Panels will work first on the Level 3 threshold because this is the point at which, when using the Initial Alternate ELPAC, the assessment will provide information to determine a student's initial classification as either EL or IFEP. This is also the point at which, when using the Summative Alternate ELPAC, it will provide information on annual student progress toward ELP and support decisions about whether a student is ready to be RFEP.

Panelists will refer to the range PLDs that describe the full range for the three levels. The threshold Level 3 student will be defined by considering what is expected of students in Level 2 compared with what is expected in Level 3 and describing what more the student just entering Level 3—the threshold Level 3 student—can do compared to the highest-performing student in Level 2.

ETS facilitators will instruct panelists on the process and will provide an exemplar in the form of a few bullet points to exemplify the level of detail expected in a threshold student definition. Panelists will be asked to review and discuss the exemplar and modify it as needed. Instructions to panelists will include a goal to limit the definitions of the threshold students to a sufficient, but not all-encompassing, description.

After the threshold Level 3 student definition is drafted and consensus is reached, panelists will complete the threshold Level 2 student definitions. Once the threshold student definitions are completed, panelists will be provided with the final definitions for use when making judgments.

## Standard Setting Methodology

Following the completion of the threshold student definitions, panelists will be trained in the judgment-making procedure. They will have an opportunity to practice making judgments before the start of actual standard setting rounds of judgments.

During training, panelists will make judgments on online forms while two ETS staff members stand by, available to help panelists as needed in the use of the online tools. A variety of items and judgment types will be presented during training. The panel facilitator will provide feedback on the panel’s judgments, and panelists will discuss the rationales used for making judgments. In this exercise, the panel facilitator will listen for rationales and comments that are consistent with the training and will redirect if comments are inconsistent with the instructions provided. After training, panelists will be asked to sign a training evaluation form confirming their understanding and readiness to proceed. Additional training will be provided as needed.

Panelists will make two rounds of judgments for each of the levels. The first round of judgments, Round 1, will be made independently, without discussion. Panelist judgments made in Round 2 will include consideration of panelists’ judgment data as well as discussions of the data and panelists’ rationales. Each panelist will enter independent Round 2 judgments at the end of the discussion, as described in the Feedback and Discussion section.

### **Item Scoring, Judgments, and Rating Scales**

The Alternate ELPAC Test Blueprint describes multiple item types, scored as one-point or two-point items. One important goal in standard setting is to reduce the cognitive complexity of making judgments. Instructions to the panelists need to be clear and understandable. The more difficult the judgment task is, the less accurate (and meaningful) the panelists’ decision is. Instructions and judgments will be more intuitive for the panelists if the ratings are aligned with the scoring rules.

For the Alternate ELPAC, the standard setting process will include both the modified Angoff judgment procedure (Brandon, 2004; Hambleton & Pitoniak, 2006) and the extended Angoff procedure (Cizek & Bunch, 2007; Hambleton & Plake, 1995). Using these two judgment types allows panelist judgments to align with the scoring of the item. Rubrics and scoring rules will be provided to panelists, along with the answer key. Panelists will be encouraged to take notes while making judgments, specifically if there are items that the panelists would like to discuss after Round 1 judgments are complete. Facilitators will ask panelists during Round 1 feedback and discussion whether they would like to discuss any of the items.

### **Modified and Extended Angoff Judgments**

For all one-point items, the modified Angoff method will be implemented. For items scored as two-point items, the extended Angoff method will be implemented. One-point items are selected-response items. Two-point items are constructed-response items scored with a rubric that indicates how a student will obtain a score of 0, 1, or 2 points. A sample modeling scoring rubric is presented in table 2, below.

Table 2. Sample Modeling Scoring Rubric

|  |  |  |
| --- | --- | --- |
| **Score of 2— Correct** | **Score of 1— Correct After Modeling** | **Score of 0— Incorrect** |
| The student provides a correct response. | The student provides a correct response after modeling. | The student provides an incorrect response. |
| Sample response: Yes. | Sample response: Yes. | Sample response: No, or not in English. |

**Modified Angoff**

The modified Angoff method is a probability-based standard setting method. For one-point items, each panelist will judge the item on the likelihood that the threshold student would answer the item correctly. Panelists will make the judgments using the following rating scale: 0, .05, .10, .20, .30, .40, .50, .60, .70, .80, .90, .95, 1.0. If an item is judged as difficult for the threshold student, the panelist would select a lower value, meaning that the probability the threshold student would answer the item correctly is low. If another item is judged as being easier for the threshold student, a higher value would be selected, indicating it is more likely that the threshold student would answer the item correctly.

Panelists will receive the final version of the threshold student definitions and will be instructed to refer to those definitions in comparison to the demands of the item for each judgment. The panelists will be advised to approach the modified Angoff judgment-making process in two stages. First, participants will consider what is the most likely probability range that the threshold student would answer the question correctly. The facilitator will encourage the panelists to consider the following rules of thumb to guide this decision:

* Items in the 0 to .30 range are those that the threshold student would have a low chance of answering correctly.
* Items in the .40 to .60 range are those that the threshold student would have a moderate chance of answering correctly.
* Items in the .70 to 1.0 range are those that the threshold student would have a high chance of answering correctly.

Second, panelists will decide how to refine their judgment within the range. For example, if a panelist thinks that there is a high chance that the threshold student would answer the question correctly, the initial decision will be in the .70 to 1 range. The second decision for the panelist will be to judge whether the likelihood of answering it correctly is .70, .80, .90, .95, or 1.

Panelists will be asked to make two judgments for each item. The overall instructions will include a reminder that, when making Level 2 (L2), followed by Level 3 (L3) judgments, it is expected that each judgment value must be at least the same as the value of the level below it—for each item. For example, if the threshold L2 judgment is .30, the threshold L3 judgment must be .30 or higher. For extended Angoff judgments, the same applies: The threshold L3 judgment must be the same as or higher than the L2 judgments. The judgments are made in an electronic file, and the software used requires that judgments are the same as or higher than the level before.

**Extended Angoff**

An extended Angoffmethod (Cizek & Bunch, 2007; Hambleton & Plake, 1995) will be used for the two-point items. For these items, a panelist will decide on the assigned score value that would most likely be earned by the threshold student for each constructed-response or two-point item. Panelists will be asked to first review the definition of the threshold student, and then to review the item, the *DFAs*, and the scoring rubric for constructed-response items. The rubric for a constructed-response item defines the quality of the evidence that would merit a response earning a particular score.

During this review, each panelist will independently consider the level of knowledge and skill required to respond to the item as well as the features of a response that would earn a particular score, as defined by the scoring rubric. Each panelist will decide on the score most likely to be earned by each threshold student from the possible values a student can earn. For example, a panelist may consider the item to be too difficult for the threshold L2 student to be able to deliver the correct answer without modeling; however, the panelist may believe that with modeling, the threshold L2 student would be able to do so. Therefore, that panelist would make the judgment that the L2 threshold score for that item would be a score of 1. Panelists will be reminded to refer to the knowledge and skills of the threshold student definition and the scoring rules and not to expect the three levels to match the three possible scores. For the two judgments—L2 and L3—the higher level should have the same or higher expectation.

## Feedback and Discussion

The purpose of feedback and discussion is to allow panelists to consider the range of judgments across the panel and hear the rationales of the other panelists. Depending on the sample sizes and based on recommendations from the CDE, panelists may receive information about student performance on some items (percent correct data), particularly in cases in which the panel’s mean judgment varies greatly from the student performance. The goal of the discussion will be for the panel to arrive at a mutual understanding of the expectations of threshold students’ performance on this test. However, there will be no expectation of a consensus rating for each item. Panelists will be encouraged to share their rationales, and the facilitator will monitor responses to ensure maximum participation by all panelists. There will be two rounds of the process of judgment, feedback, and discussion.

### **Round 1 Feedback and Discussion, and Round 2 Judgments**

Feedback and discussion from round 1 judgment data will inform round 2 judgments. As part of the post-round 1 feedback, panelists’ own judgments will be given in the context of the range of judgments across the panel, and the facilitator will share feedback on the similarities and differences of the panel judgments on the Alternate ELPAC items.

The feedback on judgments will be displayed as a summary table with the mean, minimum, and maximum judgment for each level as well as three ranges—the number of panelists rating an item from 0 to .30, the number of panelists rating an item from .40 to .60, and the number of panelists rating an item from .70 to 1.0. Displaying the feedback in this way helps panelists understand where there is agreement and is consistent with the instructions on how to think about the full range of 0 to 1.0 when making judgments.

Feedback for item judgments based on extended Angoff judgments will be displayed as 0, 1, and 2. For both types of judgments, items for which two-thirds, or 67 percent, of panelists are in the same range for a level will be considered as close to agreement. Items with a more widespread distribution of judgments will be targeted for discussion.

Panelists will be encouraged to discuss the judgments and rationales. During the discussion, they will make notes on the judgment forms and enter independent round 2 judgments for any judgments the panelists want to change at the end of the discussion.

### **Round 2 Feedback and Discussion**

After round 2 data has been analyzed, panelists will receive feedback on the round 2 judgments, based on the panelists’ judgment data. The round 2 judgment mean will be considered the final threshold score recommendation from the standard setting panel. Panelists will review round 2 feedback and will respond to a final, confidential evaluation form. Notes from the round 2 discussion will be included in the Vertical Articulation Meeting, as described in the next section.

Due to the timing of the assessment field test and the standard setting, no impact data will be available to the panelists. Impact data, or the percent classified by total score across field test forms, requires that an equating process be completed, which would allow for a comparison of total score performance by students taking different field-test forms. The equating process will take place after standard setting and before the CDE evaluates the final field test data. Thus, impact data can be incorporated into the CDE’s subsequent work, including the development of the SSPI threshold score recommendations.

### **Vertical Articulation**

Two panelists from each of the six grade-level and grade-span tests will be asked to participate in a two-hour meeting to evaluate the process and recommendations across all six pairs of threshold score judgments. The Vertical Articulation Meeting will be scheduled on March 3, after the second week of standard setting. A facilitator will guide the discussion and ask each pair of panel representatives to describe the process that transpired, including the reaching of consensus on threshold student definitions, discussions about differences in judgments, and rationales. Based on existing sample sizes as of early December 2021, the current assumption is that there may be insufficient impact data, for at least some grade levels or grade spans, available during this meeting. If by early March 2022, after reviewing the sample sizes, the CDE deems that the sample sizes are sufficient for the purpose of the vertical articulation discussion, the percentage of students classified across all grade levels and grade spans will be displayed and discussed. ETS recommends that part of the consideration in sharing this impact data include the number of student scores available, as a small sample limits confidence in interpreting this data further.

### **Final Recommendations**

The final threshold score recommendations from the standard setting panel will be the round 2 judgment results; however, any changes recommended during the Vertical Articulation Meeting will be included in the files, and the report will be shared with the CDE.

## Recommendations and Technical Report

ETS will deliver the recommended threshold scores, notes from the Vertical Articulation Meeting, and the data files containing score distributions for all grade levels and grade spans to the CDE on March 8, 2022. The sample sizes should be taken into consideration as well as the scores not having yet been equated across the two field test forms. Preliminary data files for kindergarten, grade one, and grade two will be shared with the CDE on February 28, 2022. In addition to the recommended threshold score tables designed for presentation to the SBE, the CDE may develop SSPI recommendation tables, and data analysts at ETS may verify percentages. Further discussion, before the standard setting, will be required between the CDE and ETS to define the data file requirements, composition of the tables, and a timeline for delivery.

ETS will produce and deliver the final technical report for the standard setting by April 18, 2022, or six weeks after standard setting is complete. The technical report will contain a description of the process used to set standards, a description of the panelists’ qualifications, results presented during the standard setting process, and statistical information related to the threshold score judgments (i.e., two standard errors of judgment and two standard errors of measurement above and below the panel-recommended threshold score).

## References

Brandon, P. R. 2004. “Conclusions About Frequently Studied Modified Angoff Standard-Setting Topics.” *Applied Measurement in Education.* 17: 59–88.

Cizek, G. J., and M. B. Bunch. 2007. *Standard Setting: A Guide to Establishing and Evaluating Performance Standards on Tests.* Thousand Oaks, CA: Sage Publications, Inc.

Council of Chief State School Officers (CCSSO). 2019. *English Language Proficiency Standards for English Learners with Significant Cognitive Disabilities.* Washington, DC: CCSSO. [https://ccsso.org/sites/default/files/2019-02/K-12 ELP Standards for ELs with significant cognitive disabilities 2019 02 15 .pdf](https://ccsso.org/sites/default/files/2019-02/K-12%20ELP%20Standards%20for%20ELs%20with%20significant%20cognitive%20disabilities%202019%2002%2015%20.pdf)

Hambleton, R. K., and M. J. Pitoniak. 2006. “Setting Performance Standards. In *Educational Measurement* (4th ed.), edited byR. L. Brennan. Santa Barbara, CA: Praeger Publishers.

Hambleton, R. K., and B. S. Plake. 1995. “[Using an Extended Angoff Procedure to Set Standards on Complex Performance Assessments](https://www.tandfonline.com/doi/abs/10.1207/s15324818ame0801_4)*.*” *Applied Measurement in Education.* 8(1): 41–55*.*